ELSEVIER

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Journal of Multivariate Analysis 95 (2005) 431–448

Journal of Multivariate Analysis

www.elsevier.com/locate/jmva

# A two-way analysis of variance model with positive definite interaction for homologous factors

David Causeur[a,*], Thierry Dhorne[b], Arlette Antoni[b]

[a]*Laboratoire de Mathematique Appliquees, Pole d'Enseignement Superieur et de Recherche Agronomique de Rennes, Agrocampus Rennes, CREST-ENSAI, 65, rue de St-Brieuc, 35042 Rennes cedex, France*

[b]*Laboratoire SABRES,Université de Bretagne-Sud, Rue Yves Mainguy, Vannes, France*

Received 26 September 2003

Available online 17 September 2004

**Abstract**

A special type of modelling of interaction is investigated in the framework of two-way analysis of variance models for homologous factors. Factors are said to be homologous when their levels are in a meaningful one-to-one relationship, which arise in a wide variety of contexts, as recalled by McCullagh (J. Roy. Statist. Soc. B 62 (2000) 209). The classical linear context for analysis of interaction is extended by positive definiteness restrictions on the interaction parameters. These restrictions aim to provide a spatial representation of the interaction. Properties of the maximum likelihood estimators are derived for a given dimensionality of the model. When the dimension is unknown, an alternative procedure is proposed based on a penalty approach. This approach relies heavily on random matrix theory arguments but we focus on their statistical consequences especially on the reduction of over-fitting problems in the maximum likelihood estimation. Confidence ellipses are provided for an illustrative example.



*AMS 2000 subject classification:* primary 62H15; secondary 62J10

*Keywords:* Biadditive models; Biplot; Dimensionality; Eigenvalue distribution; Homologous factors; Multidimensional scaling; Random matrix theory; Structured interaction models

* Corresponding author. Fax: +33-22348-5871.

*E-mail addresses:* david.causeur@agrorennes.educagri.fr (D. Causeur), thierry.dhorne@univ-ubs.fr (T. Dhorne), arlette.antoni@univ-ubs.fr (A. Antoni).

0047-259X/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2004.07.006

## 1. Introduction

Statistical analysis of the interaction is often based on special models that extend the classical framework of linear models by a non-linear structure of interaction: multiplicative effects, spatial effects, and so on. One of the first attempts to structure the interaction in two-way analysis of variance models is due to Tukey [16] who proposed non-additive interaction terms defined as the product of the main effects of the two factors. In the context of plant breeding, modelling of genotype-by-environment interaction motivated by Tukey [16] has led to many theoretical developments (see for instance, Mandel [11]). More recently, Denis and Gower [2] presented a review of the statistical literature on non-linear modelling of interaction and introduced the class of biadditive models as a natural extension of the multiplicative structuration of interaction.

The model presented and analyzed in the following belongs to the class of structured-interaction models. It was initially presented by Dhorne [2] in the context of factors with homologous levels. Factors are said to be homologous when their levels are the same or at least in a meaningful one-to-one relationship. Such factors are encountered, for example, in the framework of plant breeding where diallel designs are used to study parental effects on yields. McCullagh [12] provides a detailed review of the algebraic tools used for the analysis of variance with homologous factors and furthermore gives a list of applications in a wide variety of domains. Many examples of biadditive models discussed in Denis and Gower [2] are of the analysis of interaction between homologous factors: one of them was introduced by Hayman [7] in the framework of diallel models. The scope of biadditive models enables a detailed investigation of the symmetric and the anti-symmetric part of interaction. According on the objective of the analysis, both parts can give insight on the interpretation of interaction. The model presented here focuses on a symmetric modelling of interaction oriented towards a distance-based interpretation of interaction. This modelling is based on assumptions of positive definiteness on the interaction that lead to specific properties. In particular, up to random matrix theory arguments, the residual and the structural part of the model can be differentiated here to define a consistent estimator of the dimensionality.

In some of the applied contexts referred above, interaction is interpreted as a natural distance between levels of the two factors. For instance, when paired comparisons are used to compare objects in sensory experiments, the collected data present in a square table, giving the numbers of times a product, is preferred to another one by a group of panelists. As shown by Causeur and Husson [1], interaction between products can result in intransitiveness of the preference that can be modelled by introducing a latent sensory distance between the products. As another illustrative example cited by McCullagh [12], consider the geographical study of migration by square tables of numbers of people migrating from a region to another: in that case, the present modelling of interaction between the regions provides a map of these regions on which the distances reflect the intensity of inter-regions flows rather than the geographical proximity. Finally, the example that is used as a numerical illustration of our model takes place in the theory of social networks. The data set is a so-called socio matrix that reproduces the grades given by members of a group to the other ones. Introducing a kind of social distance between the members of the group in the model of interaction leads to a spatial representation of the social affinity within the group.

In Section 2, theoretical results concerning biadditive models are recalled and our Euclidean biadditive model is defined. In particular, a formal comparison with the general class of biadditive models is proposed and geometric properties of the model are described. In Section 3, maximum likelihood estimation and its properties are investigated when the dimensionality is assumed to be known. Random matrix theory arguments are used to define a penalty approach to estimate the dimensionality. This study enables a modification of the maximum-likelihood strategy, which corrects for over-fitting problems. Finally, in Section 4, the preceding results are illustrated by an example in social science.

## 2. Biadditive models

A detailed review of two-way analysis of variance models with structured interaction is provided in Denis and Gower [2]. We will focus in this section on the multiplicative biadditive models for 2 homologous factors.

**Definition 2.1.** Let $Y_{ij},\ 1\leqslant i, j\leqslant G$, denote the independently and normally distributed response variable for the $i$th level of the first and the $j$th level of the second factor, and suppose

$$E(Y_{ij}) = \mu + \alpha_i + \beta_j + \eta_{ij},\ \mathrm{Var}(Y_{ij}) = \sigma^2, \tag{1}$$

where $0 < R\leqslant G-1$ and for all $1\leqslant i, j\leqslant G$, $\eta_{ij} = \sum_{r=1}^{R}\rho_r\gamma_{ir}\gamma_{jr}$. Moreover, $\mu$, $\alpha_i,\ 1\leqslant i\leqslant G$, $\beta_j,\ 1\leqslant j\leqslant G$, $\rho_r,\ 1\leqslant r\leqslant R$ and $\gamma_{ir}$ are subject to the following identifiability restrictions:

$$\begin{array}{c} \sum_{i=1}^{G}\alpha_i = 0, \qquad \sum_{j=1}^{G}\beta_j = 0, \quad \sum_{i=1}^{G}\gamma_{ir} = 0, \text{ for all } r, \\ \sum_{i=1}^{G}\gamma_{ir}^2 = G, \text{ for all } r, \ \sum_{i=1}^{G}\gamma_{ir}\gamma_{ir'} = 0, \text{ for all } (r, r') \text{ with } r\neq r'. \end{array} \tag{2}$$

Analysis of interaction by biadditive models consists of a singular value decomposition of the $G\times G$ matrix of residuals of the additive analysis of variance sub-model.

Let $\psi_{ij} = E(Y_{ij})$. The following parametric relationships are straightforward deduced from restrictions (2):

$$\mu = \psi_{..},\ \alpha_i = \psi_{i.} - \psi_{..},\ \beta_j = \psi_{.j} - \psi_{..},$$

where $\psi_{i.} = (1/G)\sum_{j=1}^{G}\psi_{ij},\ \psi_{.j} = (1/G)\sum_{i=1}^{G}\psi_{ij},\ \psi_{..} = (1/G^2)\sum_{i=1}^{G}\sum_{j=1}^{G}\psi_{ij}$. Now, call $\Psi^{(c)}$ and $\eta^{(c)}$ the $G\times G$ matrices obtained from $\Psi = (\psi_{ij})_{1\leqslant i,j\leqslant G}$ and $\eta = (\eta_{ij})_{1\leqslant i,j\leqslant G}$, respectively, by row–column centering, then, according to Definition 2.1:

$$\eta^{(c)} = \Psi^{(c)} = \Gamma D_\rho \Gamma',$$

where $D_\rho$ is the $R\times R$ diagonal matrix containing the non-zero eigenvalues $\rho_r$ of $\eta$, $R$ is the rank of $\eta$ and $\gamma_r = (\gamma_{1r}, \gamma_{2r}, \ldots, \gamma_{Gr})'$ are the associated normalized eigenvectors.

Consequently, biadditive models allow the study of the interaction effect by the graphical tools commonly used in multivariate exploratory data analysis. For instance, diagrams based on Mandel's [11] model are shown by Kempton [10] to be attractive extensions of the performance plot usually associated to the linear two-way analysis of variance model.

Gower and Hand [6] provide a detailed review of the graphical representations helpful in interpreting interaction effects estimated throughout biadditive models.

The positive definite interaction model is now defined in the context of a complete balanced design. The impact of missing data in the square tables on the testing procedures is out of the scope of this paper. Moreover, without loss of generality, it will be assumed that each cell contains one observation.

**Definition 2.2.** Let $Y_{ij},\ 1\leqslant i, j\leqslant G$, denote the independently and normally distributed response variable for the $i$th level of the first and the $j$th level of the second factor, then assumptions (1) hold with the following modifications concerning the interaction:

$$\eta_{ij} = s\sum_{r=1}^{R}\left[\gamma_{ir}-\gamma_{jr}\right]^2,$$

where $s=\pm 1$.

The identifiability restrictions (2) are assumed for the additive parameters and the interaction parameters fulfill the following restrictions:

$$\sum_{i=1}^{G}\gamma_{ir}=0,\ \text{for all } r,\ \sum_{i=1}^{G}\gamma_{ir}\gamma_{ir'}=0,\ \text{for all } (r,r') \text{ with } r\neq r'. \tag{2'}$$

In the sequel, the $(2G+1)$-vector of additive parameters is denoted $\theta_a$ and the $RG$-vector $(\gamma_1',\ldots,\gamma_R')'$ of interaction parameters is denoted $\theta_m$.

Call $Q$ the $G\times G$ orthogonal matrix of the normalized eigenvectors of $\eta^{(c)}$ associated to the eigenvalues $(\lambda_r)_{r=1,\ldots,G}$ and $\Lambda$ the $G\times G$ diagonal matrix of the eigenvalues, then $\eta^{(c)}=\Psi^{(c)}=Q\Lambda Q'$. Moreover, under restrictions (2'), $\eta^{(c)}=\Psi^{(c)}=-2s\Gamma\Gamma'$. Equivalently, $-s\eta^{(c)}$ is positive definite with rank $R$. Therefore, if $Q_R$ stands for the $G\times R$ submatrix of $G$ obtained by keeping the eigenvectors associated to the $R$ positive (if $s=-1$, negative otherwise) eigenvalues and $\Lambda_R$ is the $R\times R$ diagonal matrix of those eigenvalues, then $\Gamma=(1/\sqrt{2})Q_R\Lambda_R^{1/2}$. The positive definiteness restriction on the modelling of interaction is in fact implicitly assumed by introducing distances between levels of the factors in the definition of interaction parameters. The relationships between the distance-based approach and the positive definiteness restriction is made clearer in the following lemma that can be deduced from Gower [5].

**Lemma 2.1.** *For $G\geqslant 1$, let $\eta$ denote a $G\times G$ matrix with generic term $\eta_{ij}$ and $\eta^{(c)}$ the row–column centered matrix with generic term $\eta_{ij}-\eta_{i.}-\eta_{.j}+\eta_{..}$, the following two propositions are equivalent*:

- *there exists $R\geqslant 1$, and $R$-vectors $\gamma_1,\ldots,\gamma_G$ such that $\eta_{ij}=(\gamma_i-\gamma_j)'(\gamma_i-\gamma_j)$,*
- *the rank of $\eta^{(c)}$ is $R$ and $-\eta^{(c)}\in\mathcal{S}^+$, where $\mathcal{S}^+$ denotes the set of positive definite symmetric matrices.*

The additive parameters of the positive interaction model can also be expressed through parameters $E(Y_{ij})=\psi_{ij}$ as follows:

$$\begin{aligned}\mu=\tfrac{1}{G}\mathrm{tr}(\Psi),\quad \alpha_i&=\tfrac{1}{2}(\psi_{i.}-\psi_{.i})-\tfrac{1}{2}\left[\tfrac{1}{G}\mathrm{tr}(\Psi)-\psi_{ii}\right],\\ \beta_j&=\tfrac{1}{2}(\psi_{.j}-\psi_{j.})-\tfrac{1}{2}\left[\tfrac{1}{G}\mathrm{tr}(\Psi)-\psi_{jj}\right],\end{aligned} \tag{3}$$

where tr(.) denotes the usual trace operator.

Note that the choice of a proper sign $s$ for the interactions depends on the latent structure of $\eta^{(c)}$: if the largest eigenvalues of $\eta^{(c)}$, in terms of absolute value, are positive, it is more interesting to consider that $s = -1$ and conversely that $s = +1$.

Let us consider a small illustrative example with $G = 4$ and such that $(2, 1, 0, -2)$ are the eigenvalues of $\eta^{(c)}$. By the biadditive modelling of interaction, it is considered that $R = 2$, which is by the way very convenient for graphical representations, and consequently the interaction model is based on the 2 largest eigenvalues 2 and $-2$. The interaction is then displayed by a two-dimensional diagram which axis are the corresponding eigenvectors. By the positive definite interaction model with $s = -1$, the two positive eigenvalues 2 and 1 will be used for modelling, which obviously leads to a different display of interaction by a diagram on which the inter-distances between the levels of the factors reflect the interaction between these levels.

Therefore, the first difference between the structures of interaction in biadditive and positive definite interaction models is the latent structure of the non-additive part of the model. In biadditive models, the eigenvectors associated to the $R$ non-zero eigenvalues are used to investigate the interaction whereas the positive definite modelling is based on the $R$ largest eigenvalues with the same sign. Moreover, the normalization of the eigenvectors in biadditive and positive definite interaction models are different. In fact, by positive definite interaction models, this normalization is consistent with a spatial representation of the levels of the factors connecting interaction with distances between levels. This graphical display of the interaction relies on multidimensional scaling techniques introduced by Gower [5].

As also discussed later, when $R$ is not a rank, as in biadditive models, but a number of positive eigenvalues, symmetry of the random error can be used in estimation procedures of $R$ in order to decrease the bias of the maximum likelihood estimator.

## 3. Maximum likelihood estimation

In the following proposition, maximum likelihood estimators of the interaction parameters are derived under the assumptions introduced in Definition 2.2. Due to the functional invariance of the maximum likelihood method, the estimators of the additive parameters are straightforwardly deduced from the relationships (3).

**Proposition 3.1.** *For $G \geqslant 1$, let $\mathbf{Y}$ denote the $G \times G$ matrix with generic term $Y_{ij}$, where the $Y_{ij}$ are introduced in Definition 2.2. Let $\mathbf{Y}^{(c)}$ denote the $G \times G$ matrix obtained by row–column centering of $\mathbf{Y}$. Let $\mathbf{Y}_s^{(c)}$ denote the symmetric matrix $(\mathbf{Y}^{(c)} + \mathbf{Y}^{(c)'})/2$. The singular value decomposition of $\mathbf{Y}_s^{(c)}$ is defined by $\mathbf{Y}_s^{(c)} = \hat{Q}\hat{\Lambda}\hat{Q}'$, where $\hat{Q}$ is the $G \times G$ orthogonal matrix of normalized eigenvectors and $\hat{\Lambda}$ is the $G \times G$ diagonal matrix of the associated eigenvalues $(\hat{\lambda}_r)_{1 \leqslant r \leqslant G}$ in decreasing order. Now, let $p$ denote the number of positive, if $s = +1$, or negative, if $s = -1$, eigenvalues of $\mathbf{Y}_s^{(c)}$. Call $\hat{\Lambda}_p$ the $p \times p$ diagonal matrix of those eigenvalues in decreasing order and $\hat{Q}_p$ the $G \times p$ submatrix of $\hat{Q}$ of the associated eigenvectors.*

*The maximum likelihood estimator $\hat{R}$ of $R$ is given by $\hat{R} = p$.*

*The maximum likelihood estimator* $\hat{\mathbf{\Gamma}}$ *of* $\mathbf{\Gamma}$ *is defined as follows:*

$$\hat{\mathbf{\Gamma}} = \begin{cases} \frac{1}{\sqrt{2}}\hat{Q}_p\hat{\Lambda}_p^{\frac{1}{2}}, & \text{if } p \geqslant 1, \\ 0, & \text{if } p = 0. \end{cases} \tag{4}$$

*The maximum likelihood estimator* $\hat{\eta}_{ij}$ *of* $\eta_{ij}$ *is:*

$$\hat{\eta}_{ij} = -(\hat{Q}_p\hat{\Lambda}_p\hat{Q}'_p)_{ij} + \frac{1}{2}(\hat{Q}\hat{\Lambda}_p\hat{Q}'_p)_{ii} + \frac{1}{2}(\hat{Q}_p\hat{\Lambda}_p\hat{Q}'_p)_{jj}.$$

*Finally:*

$$\hat{\sigma}^2 = \frac{1}{G^2}\mathrm{tr}\left(\left[\mathbf{Y}^{(c)} - \mathbf{Y}_s^{(c)}\right]\left[\mathbf{Y}^{(c)} - \mathbf{Y}_s^{(c)}\right]\right) + \frac{1}{G^2}\sum_{\hat{\lambda}_i \notin \hat{\Lambda}_R}\hat{\lambda}_i^2.$$

**Proof.** $\hat{\eta}$ is defined as follows:

$$\hat{\eta} = \arg\min_{\eta\in\Omega}\min_{\theta_a\in\Theta_a}\sum_{i=1}^{G}\sum_{j=1}^{G}(Y_{ij} - \mu - \alpha_i - \beta_j - \eta_{ij})^2,$$

where $\Theta_a$ is the set of $(2G+1)$-vectors $\theta_a$ under the usual sum-to-zero restrictions and $\Omega$ denotes the set of $G \times G$ matrix $\eta$ under restrictions (2').

Minimizing the former sum of squares relative to $\theta_a$ leads to the following expression:

$$\hat{\eta} = \arg\min_{\eta\in\Omega}\sum_{i=1}^{G}\sum_{j=1}^{G}\left(\left[Y_{ij\cdot} - \eta_{ij}\right] - \left[Y_{i\cdot\cdot} - \eta_{i\cdot}\right] - \left[Y_{\cdot j\cdot} - \eta_{\cdot j}\right] + \left[Y_{\cdots} - \eta_{\cdot\cdot}\right]\right)^2,$$

or, equivalently:

$$\hat{\eta} = \arg\min_{\eta\in\Omega}\mathrm{tr}\left[\mathbf{Y}^{(c)} - \eta^{(c)}\right]\left[\mathbf{Y}^{(c)} - \eta^{(c)}\right]',$$

where $\mathbf{Y}^{(c)}$ and $\eta^{(c)}$ are derived from $\mathbf{Y}$ and $\eta$ throughout row–column centering.

It follows from Lemma 4.1:

$$\hat{\eta} = \arg\min_{\mathbf{S}=-s\eta^{(c)}\in\mathcal{S}^+}\mathrm{tr}\left[-\mathbf{Y}^{(c)} - \mathbf{S}\right]\left[-\mathbf{Y}^{(c)} - \mathbf{S}\right]'.$$

Therefore, $\hat{\eta}$ can be deduced from the singular value of $-\mathbf{Y}_s^{(c)}$ by keeping only the positive (if $s = -1$) or negative (if $s = +1$) eigenvalues and the associated eigenvectors. Other estimators are straightforwardly deduced from $\hat{\eta}$.

Note that the residual sum of squares of the former minimization is split into two parts: $\mathrm{tr}([\mathbf{Y}^{(c)} - \mathbf{Y}_s^{(c)}][\mathbf{Y}^{(c)} - \mathbf{Y}_s^{(c)}])$ which is the residual part from the projection of the table $\mathbf{Y}^{(c)}$ onto the set of symmetric matrices and $\sum_{\hat{\lambda}_i \notin \hat{\Lambda}_R}\hat{\lambda}_i^2$ which is the residual part from the projection of the symmetrized table $-s\mathbf{Y}_s^{(c)}$ onto the set of positive definite matrices. The formula for $\hat{\sigma}^2$ is deduced from this residual sum of squares by dividing by the number of observations in the table. □

Table 1
Decomposition of the total sum of squares of the positive interaction model

| Effect | Sum of squares | | Degree of freedom |
|---|---|---|---|
| Total | $\sum_{k=1}^{K}\sum_{i=1}^{G}\sum_{j=1}^{G}(Y_{ijk}-Y_{...})^2$ | | $KG^2-1$ |
| 1st factor | $KG\sum_{i=1}^{G}(Y_{i..}-Y_{...})^2$ | | $G-1$ |
| 2nd factor | $KG\sum_{j=1}^{G}(Y_{.j.}-Y_{...})^2$ | | $G-1$ |
| Interaction | Globally | $K\sum_{i=1}^{p}\lambda_i^2$ | $RG-\frac{R(R+1)}{2}$ |
| | Dim. 1 | $K\lambda_1^2$ | $G-1$ |
| | Dim. 2 | $K\lambda_2^2$ | $G-2$ |
| | Dim. 3 | $K\lambda_3^2$ | $G-3$ |
| | $\vdots$ | $\vdots$ | $\vdots$ |
| | Dim. $R$ | $K\lambda_R^2$ | $G-R$ |
| Residual | $\frac{K}{4}\mathrm{tr}\left(\left[\mathbf{Y}^{(c)}-\mathbf{Y}_s^{(c)}\right]\left[\mathbf{Y}^{(c)}-\mathbf{Y}_s^{(c)}\right]\right)$ $+\sum_{k=1}^{K}\sum_{i=1}^{G}\sum_{j=1}^{G}\left(Y_{ijk}-Y_{ij.}\right)^2$ $+K\sum_{\hat{\lambda}_i\notin\hat{\Lambda}_R}\hat{\lambda}_i^2$ | | $KG^2-RG-2G$ $+\frac{R(R+1)}{2}-2$ |

In the context of a complete balanced design with $K$ replicates per cell, the estimates of the parameters of the positive interaction model are deduced from proposition 3.1 by replacing $Y_{ij}$ by the mean value of $Y$ in the cell $(i, j)$. In that case, the intra-cell variability is added to the estimate of the residual variance:

$$\hat{\sigma}^2 = \frac{1}{G^2}\mathrm{tr}\left(\left[\mathbf{Y}^{(c)}-\mathbf{Y}_s^{(c)}\right]\left[\mathbf{Y}^{(c)}-\mathbf{Y}_s^{(c)}\right]\right)+\frac{1}{G^2}\sum_{\hat{\lambda}_i\notin\hat{\Lambda}_R}\hat{\lambda}_i^2 + \frac{1}{KG^2}\sum_{i=1}^{G}\sum_{j=1}^{G}\sum_{k=1}^{K}\left(Y_{ijk}-Y_{ij.}\right)^2,$$

The previous expression of $\hat{\sigma}^2$ relies on a natural decomposition of the total sum of squares into sums of squares related to the additive sub-model, the interaction effect and the residual term. Moreover, note that the interaction models are nested which allow for a sequential decomposition of the interaction sum of squares into parts related to the dimensions. This decomposition is displayed in Table 1 and the corresponding degrees of freedom are provided. By analogy with the usual procedures used in the linear context, this table can be a basis for the calculation of F-statistics aiming at testing the effects in the model.

In biadditive models, the dimensionality $R$ introduced in Definition 2.1 is assumed to be a known parameter. However, in practice, the choice of $R$ is mainly motivated by an optimal representation of interaction. In that context, the traditional empirical methods used in multivariate exploratory data analysis are helpful to choose a suitable value for $R$. As mentioned above, these empirical approaches may be confirmed by testing procedures based on a decomposition of the sum of squares related to the interaction effect in Table 1. An alternative penalty approach is investigated in the next section. Conversely, in

the model described in Definition 2.2, $R$ is assumed to be unknown and maximum likelihood estimation provides an objective criterion for the choice of $R$.

## 4. Asymptotic properties

As encountered for any reduced rank model, asymptotic theory usually invoked in the multinormal context is not suited in the case of the positive interaction model due to a lack of prior knowledge on the dimensionality of the model: the number of parameters in this model depends on $R$, which is itself an unknown parameter. Therefore, in this context, asymptotic properties are studied in the following throughout simulations and they are formally achieved only in the case of a prior value for $R$.

From now on, it is assumed for convenience that $s = +1$.

### 4.1. Known value of the dimensionality R

In this framework, as proposed by Denis and Gower [2], Silvey's [14] results are helpful to derive the properties of the maximum likelihood estimates under quite general restrictions.

**Proposition 4.1.** *Let $\theta = (\theta_a', \theta_m')'$ stand for the $(1+2G+RG)\times 1$ vector of parameters in the model described in Definition 2.2 with $K$, $K \in \mathbf{N}^*$, replicates per cell, let $\hat{\theta}$ stand for the maximum likelihood estimator of $\theta$ deduced from Proposition 3.1, then, for large values of $K$:*

$$\sqrt{K}(\hat{\theta} - \theta) \sim \mathcal{N}(0, \sigma^2 V_\theta),$$

*where $V_\theta$ is the upper left $(RG+2G+1)\times(RG+2G+1)$ block in the following matrix:*

$$\begin{pmatrix} M(\theta) & L(\theta) \\ L'(\theta) & 0 \end{pmatrix}^{-1},$$

*and*

- *$M(\theta) = J'(\theta)J(\theta)$ and $J(\theta)$ denote the $(KG^2)\times(RG+2G+1)$ partitioned matrix*

$$J(\theta) = \begin{pmatrix} J(\theta_a) & J(\theta_m) \end{pmatrix},$$

*$J(\theta_a)$ is the usual $(KG^2)\times(1+2G)$ design matrix of the additive two-way analysis of variance sub-model and $J(\theta_m) = J^{(1)}(\theta_m) + J^{(2)}(\theta_m)$, $J^{(1)}(\theta_m)$ is the $(KG^2)\times(RG)$ block-diagonal matrix which ith $(KG)\times R$ block $J_i^{(1)}(\theta_m)$ is defined as follows:*

$$J_i^{(1)}(\theta_m) = 2\,\mathbf{1}_K \otimes \left[\gamma_i' \otimes \mathbf{1}_G - \Gamma'\right]$$

*and* $J^{(2)}(\theta_m)$ *is the* $(KG^2) \times (RG)$ *following partitioned matrix*:

$$J^{(2)}(\theta_m) = \begin{pmatrix} J_1^{(2)}(\theta_m) \\ J_2^{(2)}(\theta_m) \\ \vdots \\ J_G^{(2)}(\theta_m) \end{pmatrix},$$

*for all* $j = 1, \ldots, G$, $J_j^{(2)}(\theta_m)$ *is the* $(KG) \times (RG)$ *block-diagonal matrix which ith* $K \times R$ *block is* $2[\gamma_i' - \gamma_j'] \otimes \mathbf{1}_K$.

- $L(\theta)$ *denotes the following* $(2 + \frac{R(R+1)}{2}) \times (1 + 2G + RG)$ *matrix*:

$$L(\theta) = \begin{pmatrix} L_a(\theta_a) & 0 \\ 0 & L_m(\theta_m) \end{pmatrix},$$

*where* $L_a(\theta_a)$ *stands for the following* $2 \times (1 + 2G)$ *matrix*:

$$L_a(\theta_a) = \begin{pmatrix} 0 & \mathbf{1}_G' & 0 \\ 0 & 0 & \mathbf{1}_G' \end{pmatrix}$$

*and* $L_m(\theta_m)$ *denotes the following* $\frac{R(R+1)}{2} \times (RG)$ *matrix*:

$$L_m(\theta_m) = \begin{pmatrix} I_R & I_R & \ldots & I_R \\ L_m^{(1)}(\theta_m) & L_m^{(2)}(\theta_m) & \ldots & L_m^{(G)}(\theta_m) \end{pmatrix},$$

$L_m^{(j)}(\theta_m)$, $j = 1, \ldots, G$ *are* $\frac{R(R-1)}{2} \times R$ *matrix partitioned as follows*:

$$L_m^{(j)}(\theta_m) = \begin{pmatrix} L_m^{(1j)}(\theta_m) \\ L_m^{(2j)}(\theta_m) \\ \vdots \\ L_m^{(q-1,j)}(\theta_m) \end{pmatrix},$$

$L_m^{(ij)}(\theta_m)$ *is a* $(R - i) \times R$

$$L_m^{(ij)}(\theta_m) = \begin{pmatrix} \gamma_{j,i+1} & \gamma_{ji} & 0 & \ldots & 0 \\ \gamma_{j,i+2} & 0 & \gamma_{ji} & \ldots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ \gamma_{jR} & 0 & 0 & \ldots & \gamma_{ji} \end{pmatrix}.$$

**Proof.** Asymptotic normality of $\hat{\theta}$ is deduced from Silvey's [14] results concerning maximum likelihood estimation in multinormal models under restrictions on parameters. The asymptotic variance is achieved throughout derivatives of $J(\theta)$ and $L(\theta)$, namely the derivatives relative to $\theta$ of the $(KG^2)$-valued expectation vector and the $(2 + \frac{R(R+1)}{2})$-valued restriction function, respectively. □

Due to the restrictions on the parameters, the parametric dimension of the positive interaction model is $q = 1 + 2G + RG - (2 + \frac{R(R+1)}{2})$, namely $1 + 2G - 2$ for the additive parameters and $RG - \frac{R(R+1)}{2}$ for the interaction parameters. Therefore, $q$ is also the rank of the asymptotic variance matrix of the maximum likelihood estimates.

Asymptotic variances for the estimators of the interaction parameters are easily deduced from Proposition 4.1.

**Proposition 4.2.** *Let $\eta^*$ stand for the $G^2$-vector obtained by collecting the column vectors of $\eta$ and $\hat{\eta}^*$ the maximum likelihood estimator of $\eta^*$ deduced from Proposition 3.1, then $\hat{\eta}^*$ is asymptotically unbiased for $\eta^*$ and asymptotic variance of $\hat{\eta}^*$ is given by*:

$$\mathrm{Var}(\hat{\eta}^*) = \frac{\sigma^2}{K} J'_\eta(\theta_m) V_{\theta_m} J_\eta(\theta_m),$$

*where $V_{\theta_m}$ denotes the $RG \times RG$ lower right block in the matrix $V_\theta$ derived in proposition 4.1 and*

$$J_\eta(\theta_m) = J^{(1)}_\eta(\theta_m) + J^{(2)}_\eta(\theta_m),$$

*where $J^{(1)}_\eta(\theta_m)$ is the $G^2 \times RG$ block-diagonal matrix which ith $G \times R$ block $J^{(1)}_{\eta,i}(\theta_m)$ is defined as follows*:

$$J^{(1)}_{\eta,i}(\theta_m) = 2 \otimes \left[\gamma'_i \otimes \mathbf{1}_G - \Gamma'\right]$$

*and $J^{(2)}_\eta(\theta_m)$ is the $G^2 \times (RG)$ following partitioned matrix*:

$$J^{(2)}_\eta(\theta_m) = \begin{pmatrix} J^{(2)}_{\eta,1}(\theta_m) \\ J^{(2)}_{\eta,2}(\theta_m) \\ \vdots \\ J^{(2)}_{\eta,G}(\theta_m) \end{pmatrix},$$

*for all $j = 1, \ldots, G$, $J^{(2)}_{\eta,j}(\theta_m)$ is the $G \times (RG)$ block-diagonal matrix which ith $1 \times R$ block is $2[\gamma'_i - \gamma'_j]$.*

**Proof.** Let $\frac{\partial \eta^*}{\partial \theta_m}$ denote the $G^2 \times (RG)$ matrix containing derivatives of $\eta_{ij}$, $i, j = 1, \ldots, G$, relative to $\theta_m$ then, asymptotic variance of $\hat{\eta}^*$ is derived throughout the asymptotic variance of $\hat{\theta}_m$ according to:

$$\mathrm{Var}(\hat{\eta}^*) = \frac{\partial \eta^*}{\partial \theta_m}' \mathrm{Var}(\theta_m) \frac{\partial \eta^*}{\partial \theta_m}.$$

Expression (5) is finally obtained by showing that $\frac{\partial \eta^*}{\partial \theta_m} = J_\eta(\theta_m)$. □

In the following, this asymptotic variance matrix will allow the derivations of elliptical confidence regions for the parameters $\Gamma$.

### 4.2. Unknown value of the dimensionality R

As mentioned in Proposition 3.1, the dimensionality $R$ is estimated by the number of positive eigenvalues of $-Y_s^{(c)}$, or equivalently of $-\eta^{(c)} - (\varepsilon^{(c)} + \varepsilon^{(c)'})/2$, where $\varepsilon^{(c)}$ is obtained from $\varepsilon = (\varepsilon_{ij})_{1 \leqslant i,j \leqslant G}$ by row–column centering. It can also be deduced from the singular value decomposition of $-\eta^{(c)}$ that $\hat{R}$ is the number of positive eigenvalues of the randomly perturbed matrix $\Lambda_\varepsilon = \Lambda - (Q'\varepsilon^{(c)}Q + Q'\varepsilon^{(c)'}Q)/2$, where $Q'\varepsilon^{(c)}Q$ is a matrix which entries are independently distributed according to a normal distribution with mean 0 and standard deviation $\sigma$ or $\sigma/\sqrt{K}$ in the case of $K$ replications per cell.

The study of the distribution of the eigenvalues of random matrices is quite usual in some special fields of applications such as nuclear physics (see Edelman [3] for a review of the main results). In the random matrix theory, the symmetric random perturbation involved in our problem is called Gaussian Orthogonal Ensemble (GOE) (see for instance Mehta [13]). Although there is a wide literature on the distribution of the eigenvalues of a GOE, most of the results are stated in the null case where $\Lambda = 0$. However, some attempts to apply those results to the dimensionality problem in multivariate statistics are due to Johnstone [9] and Hoyle and Rattray [8]. As mentioned by these authors, the eigenvalues of $\Lambda_\varepsilon$ are much more spread out that the diagonal values of $\Lambda$, which can considerably complicate the estimation of the dimensionality. In order to make it more concrete, 5000 simulated $12 \times 12$ data matrices $\Lambda_\varepsilon$ have been drawn with $\sigma/\sqrt{K} = 1$ and only 2 non-zero eigenvalues in $\Lambda$, namely: $(\lambda_1 = 12, \lambda_2 = 6)$, $(\lambda_1 = 20, \lambda_2 = 10)$ and $(\lambda_1 = 30,\ \lambda_2 = 15)$. The distributions of the mean eigenvalues of $\Lambda_\varepsilon$ are plotted on Fig. 1 both for $\Lambda \neq 0$ and for $\Lambda = 0$.

First, the graphs illustrate the fact that the number $\hat{R}$ of positive mean eigenvalues of $\Lambda_\varepsilon$ is the same whenever $\Lambda = 0$ or not. This explains that $\hat{R}$ cannot be a consistent estimator of $R$.

Moreover, for small values of $\lambda$, it is hardly impossible to detect any changes in the distribution of the mean eigenvalues that could give insight on the true number of non-zero eigenvalues.

For large $G$ and $K$, approximated values of the mean eigenvalues of a GOE can be deduced from Johnstone's [9] asymptotic results for Wishart matrices. The following result can therefore be seen as a modified version of Johnstone [9] convergence theorem dedicated to GOE rather than Wishart matrices.

$$\mu_G = 2G\left(1 + \sqrt{\frac{G-1}{G}}\right)^2,$$

$$\sigma_G = 2(\sqrt{G-1} + \sqrt{G})\left(\frac{1}{\sqrt{G-1}} + \frac{1}{\sqrt{G}}\right)^{\frac{1}{3}}.$$

If $l_1 > l_2 > \cdots > l_G$ stand for the eigenvalues of $\Lambda_\varepsilon$, in the case $\Lambda = 0$, then:

$$\frac{(\sqrt{K}l_1/\sigma)^2 - \mu_G}{\sigma_G} \sim F_1,$$

where $F_1$ is the so-called Tracy–Widom law of order 1, which distribution function is given by Tracy and Widom [15] as the solution of a nonlinear differential equation.

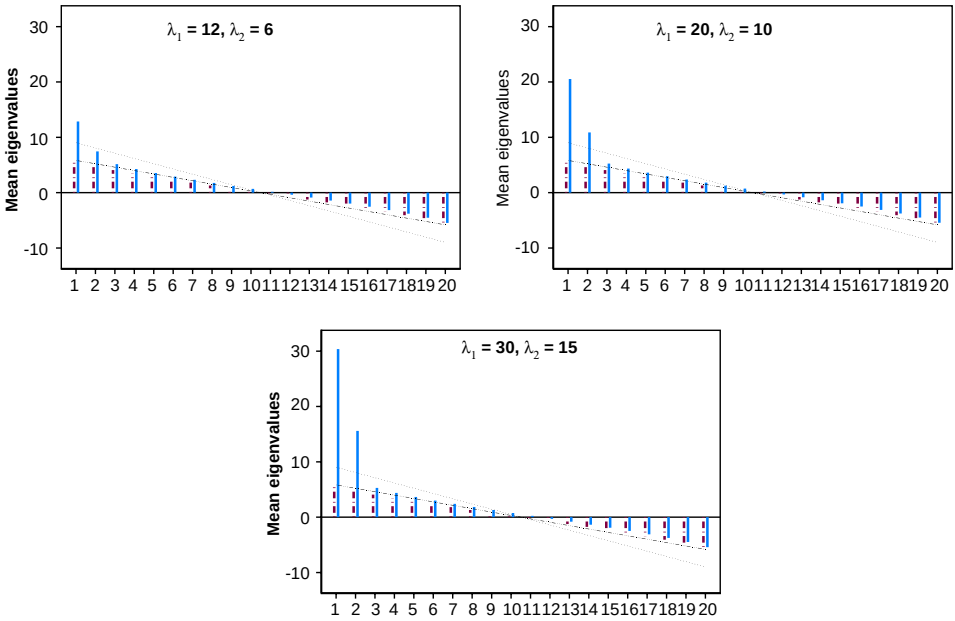


Fig. 1. Distributions of the mean eigenvalues of $\Lambda_\varepsilon$. Solid bars are for $\Lambda \neq 0$ and dotted bars for $\Lambda = 0$. The dotted and the dashed lines are, respectively, approximations for the mean and the 90% percentile eigenvalues.

Approximations for large matrices of the expectation and the 90% percentile $q_{0.90}(l_1)$ of the largest eigenvalue of a GOE are deduced by a first order Taylor expansion of the square root function:

$$\mathbb{E}(l_1) = \frac{\sigma}{\sqrt{K}}\left[\mu_G + \sigma_G \mathbb{E}(F_1)\right]^{\frac{1}{2}} - \frac{\sigma}{4\sqrt{K}}\left[\mu_G + \sigma_G \mathbb{E}(F_1)\right]^{-\frac{3}{2}} \sigma_G^2 \operatorname{Var}(F_1),$$

$$q_{0.90}(l_1) = \frac{\sigma}{\sqrt{K}}\left[\mu_G + \sigma_G q_{0.90}(F_1)\right]^{\frac{1}{2}},$$

where, according to Tracy and Widom [15], $\mathbb{E}(F_1) \approx -1.21$, $\operatorname{Var}(F_1) \approx 1.27$ and $q_{0.90}(F_1) \approx 0.45$. An empirical method is now described to obtain approximated values of the other mean eigenvalues and the associated 90% percentiles. First, note that the non-zero eigenvalues of $\Lambda_\varepsilon$ for $\Lambda \neq 0$ seem to pull up the other ones. This can be seen, at least when the non-zero eigenvalues are large enough, on the graphs of Fig. 1: the two largest eigenvalues of $\Lambda_\varepsilon$ are markedly larger for $\Lambda \neq 0$ than the two largest eigenvalues of the GOE whereas the other ones are only slightly larger. Moreover, due to the symmetry in the distribution of the mean eigenvalues, $\mathbb{E}(l_G) \approx -\mathbb{E}(l_1)$. The decrease in the mean eigenvalues can be well approximated by a line that joins the mean of the largest positive eigenvalue to the mean of the largest negative eigenvalue. This line is plotted on the graphs of Fig. 1 together with the line that joins the 90% percentiles. It follows:

$$\mathbb{E}(l_i) \approx \mathbb{E}(l_1)\left[1 - 2\frac{i-1}{G-1}\right],$$

Table 2
Mean values of $\tilde{R}$ over 5000 simulations with $R = 2$ and for various values of $G$ and $(\lambda_1, \lambda_2 = \lambda_1/2)$

| $G$ | $\lambda_1$ | | | |
|---|---|---|---|---|
| | 6 | 12 | 20 | 30 |
| 12 | 0.538 | 2.023 | 2.246 | 2.275 |
| 20 | 0.055 | 1.228 | 1.999 | 2.002 |
| 30 | 0.002 | 0.998 | 1.874 | 2.000 |
| 50 | 0.000 | 0.453 | 1.081 | 1.999 |
| 80 | 0.000 | 0.002 | 1.000 | 1.513 |

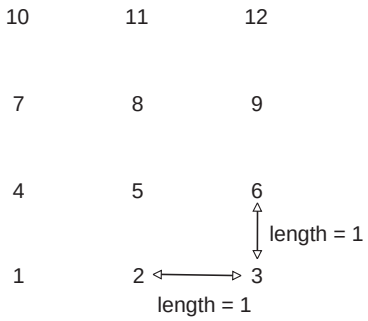


Fig. 2. Interaction parameters in the simulation feature are squared inter-distances between points.

$$q_{0.90}(l_i) \approx q_{0.90}(l_1)\left[1 - 2\frac{i-1}{G-1}\right]. \tag{5}$$

As a first consequence, the previous result gives an idea of the conditions on the non-zero eigenvalues $\lambda_i$ for a good identification of the structurally positive eigenvalues of the perturbed matrix $\Lambda_\varepsilon$. Indeed, if $\lambda_i < q_{0.90}(l_i)$, the $i$th eigenvalue of $\Lambda_\varepsilon$ is not markedly different of that of a GOE, which makes the estimation of $R$ very difficult. For instance, in our simulation feature, for $(\lambda_1 = 12, \lambda_2 = 6)$, $q_{0.90}(l_2) \approx 9.7$, which results in a bad separation between $\lambda_2$ and the 2nd eigenvalue of the GOE.

According to the previous calculations and especially to expressions (5), the following estimator $\tilde{R}$ of $R$ is proposed:

$$\tilde{R} = \#\left\{\hat{\lambda}_i, i = 1, \ldots, G\ \hat{\lambda}_i \geqslant q_{0.90}(l_i)\right\}.$$

Table 2 reproduces the mean value of $\tilde{R}$ over 5000 simulations in which $R = 2$, $\sigma/\sqrt{K} = 1$ and $(G, \lambda_1, \lambda_2 = \lambda_1/2)$ is chosen to cover a wide variety of situations. These results confirm that, when the eigenvalues $\lambda_1$ or $\lambda_2$ are markedly larger than $\sqrt{G}$, the estimation of $R$ by $\tilde{R}$ is relatively good. Otherwise, $\tilde{R}$ under-estimates $R$.

In order to show the impact of a more accurate estimation of the dimensionality on the estimation of the interaction parameters, 5000 data tables are simulated that consists in the randomly perturbed squared distances between 12 points as plotted on Fig. 2.

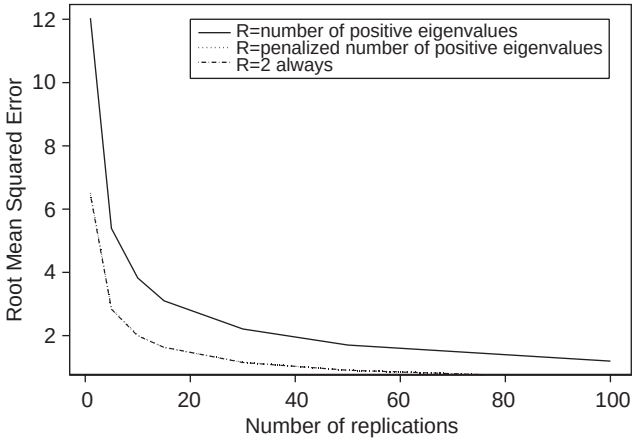


Fig. 3. RMSE for three strategies of estimation of the squared distances between the points of Fig. 2.

Table 3
RMSE for three strategies of estimation of the squared distances between the points of Fig. 2

| $K$ | Estimation of $R$ | | |
|---|---|---|---|
| | $\hat{R}$ | $\tilde{R}$ | $R=2$ |
| 1 | 12.03 | 6.51 | 6.43 |
| 5 | 5.39 | 2.86 | 2.84 |
| 10 | 3.82 | 2.01 | 1.99 |
| 15 | 3.10 | 1.62 | 1.62 |
| 30 | 2.21 | 1.16 | 1.15 |
| 50 | 1.70 | 0.90 | 0.90 |
| 100 | 1.19 | 0.63 | 0.63 |

The perturbations at each cell are independent normal with mean 0 and standard deviation $1/\sqrt{K}$. Note that, in that situation, $-\eta^{(c)}$ has two positive eigenvalues $\lambda_1 = 30$ and $\lambda_2 = 16$. The accuracy criterion used for this comparison is the usual root mean squared error (RMSE):

$$\left(\sum_{i=1}^{G}\sum_{j=1}^{G}\mathbb{E}\left(\hat{\eta}_{ij}-\eta_{ij}\right)^2\right)^{1/2},$$

which is estimated by averaging over the 5000 simulations. Table 3 gives the RMSE for each of three strategies based on $\hat{R}$, $\tilde{R}$ and a known value $R = 2$. Fig. 3 shows the positive impact of replacing $\hat{R}$ by $\tilde{R}$ on the accuracy of estimation. In fact, the estimation strategy based on $\tilde{R}$ seems to be as accurate as if the dimensionality were known to be $R = 2$ (Fig. 3, Table 3).

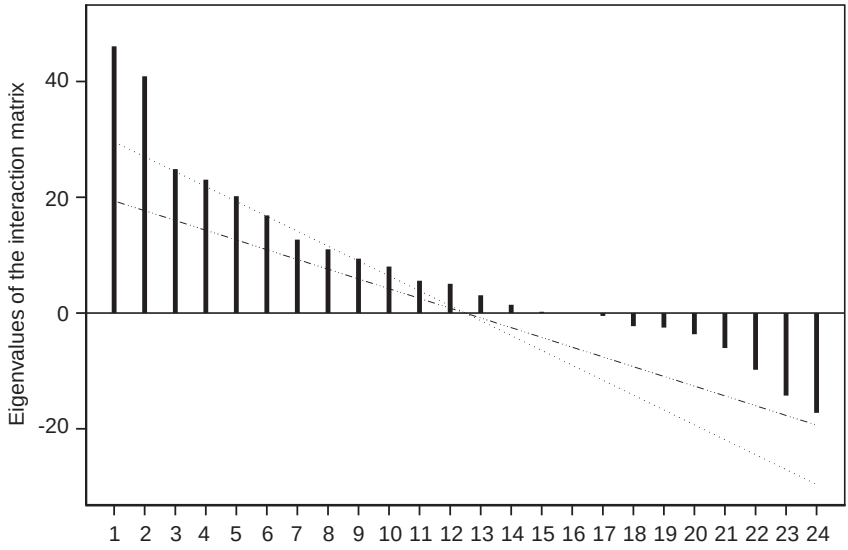


Fig. 4. Eigenvalues of the interaction matrix for the socio-matrix example. The dotted and the dashed lines are, respectively, approximations for the mean and the 90% percentile eigenvalues of the random GOE perturbation matrix.

## 5. Illustrative example

One famous method used in social network analysis and research in group dynamics is the use of the so-called socio matrices. Such matrices are used to model the relationships between the group members. The square data table reproduced in de Falguerolles and Van der Heijden [4] is an example of socio matrix used by the authors to present and compare various biadditive models. This data set contains the grades given by each of the 24 pupils of a class (in rows) to all other schoolmates (in columns). The grades are integer values ranging from 0 to 20. In the original data set, the diagonal entries are missing but, following de Falguerolles and Van der Heijden [4], they are set to 20. Our Euclidean biadditive model will be used here to model the interaction between the pupils through a social distance structure between those pupils.

First, we will focus on the estimation of the dimensionality $R$ to be used in the interaction part of the model. Fig. 4 represents the distribution of the eigenvalues of the interaction matrix. First, this figure shows that the positive eigenvalues are larger than the negative ones, which leads to choose $s = -1$ in the model. This means that the interaction terms will modify the mean affinity of a pupil for another, derived from the additive sub-model, by decreasing it as much as those pupils are socially distant.

Our penalty approach for estimating the dimensionality is based on a preliminary estimation of the residual standard deviation $\sigma$. Table 4 gives the df-corrected estimated value of $\sigma$ for plausible values of $R$. As these estimated values do not vary much from 3, this value will be used as an input in the calculation of $\tilde{R}$.

Table 4
Estimated values of $\sigma$ for different values of $R$

| $R$ | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|
| $\hat{\sigma}$ | 3.13 | 2.99 | 2.85 | 2.75 | 2.68 | 2.66 | 2.65 | 2.66 | 2.68 |

Table 5
Sequential analysis of variance table

| Effect | Sum of squares | Degree of freedom | Mean squares | $F$ | $p$-value |
|---|---|---|---|---|---|
| Row factor | 1307.832 | 23 | 56.86 | 5.80 | 0.00 |
| Column factor | 2011.998 | 23 | 87.48 | 8.92 | 0.00 |
| Interaction | | | | | |
| Dim. 1 | 2122.71 | 23 | 92.29 | 9.41 | 0.00 |
| Dim. 2 | 1672.132 | 22 | 76.01 | 7.75 | 0.00 |
| Residual | 4758.62 | 485 | 9.81 | | |

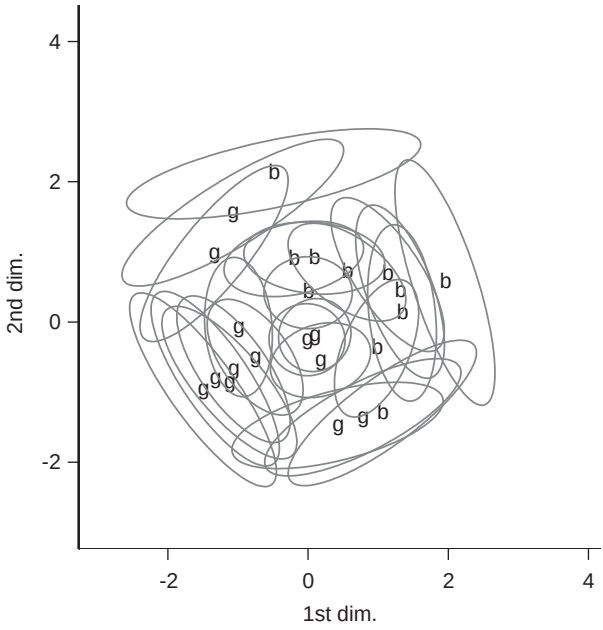


Fig. 5. Biplot of the interaction. Boys are denoted by b and girls by g.

The expected mean and percentile lines for the GOE random perturbation are represented on the graph in Fig. 4. Although the percentile line does not separate very clearly a bulk of larger eigenvalues from others, we propose to consider that $R = 2$.

The analysis of variance Table 5 confirms the significance of the biadditive Euclidean model with $R = 2$.

On the basis of the asymptotic variances derived in Section 4, the biplot of the interaction and confidence ellipses with level 0.90 are provided in Fig. 5. This biplot reflects a marked separation between boys and girls but the confidence ellipses within genders are partially overlapping, which suggests no further sub-groups structure.

## 6. Conclusion

The present paper aims at proposing a model for interaction between homologous factors. This model is particularly interesting when interaction can be beneficially interpreted with regard to a latent spatial structure. Testing procedures are proposed for the effects of each dimension in the interaction. Another strategy, based on the eigenvalues of the interaction matrix, is proposed to correct for the over-fitting problem of the maximum likelihood. This approach is essentially possible due to the positive definiteness of the interaction matrix.

Analysis of interaction is completed by a diagram with confidence ellipses that help identify groups of levels with high interaction. This diagram gives more insight in the individual level-by-level interaction parameters.

A current axis for improving and extending the present model is the development of maximum likelihood estimation procedures with missing data. In the framework of the analysis of the flows of migration between regions, situations of incomplete datasets are frequently encountered due to the large numbers of regions that are accounted for in these studies. In that case, the positive definiteness of the interaction could be help for avoiding time-consuming EM strategies.

## References

[1] D. Causeur, F. Husson, A 2-dimensional extension of the Bradley-Terry model for paired comparisons, J. Statist. Plann. Inference, 2004, to appear.
[2] J.B. Denis, J.C. Gower, Asymptotic confidence regions for biadditive models: interpreting genotype×environment interactions, Appl. Statist. 45 (4) (1996) 479–493.
[3] A. Edelman, The probability that a random real gaussian matrix has $k$ real eigenvalues, related distributions, and the circular law, J. Multivariate Anal. 60 (1997) 203–232.
[4] A. de Falguerolles, P.G.M. Van der Heijden, Reduced-rank quasi symmetry and quasi skew-symmetry: a generalized bi-linear model approach, Ann. Fac. Sci. Toulouse. XI (4) (2002) 507–524.
[5] J.C. Gower, Some distance properties of latent root and vector methods used in multivariate analysis, Biometrika 53.3 and 4 (1966) 325–338.
[6] J.C. Gower, D.J. Hand, Biplots, Monographs on Statistics and Applied Probability, vol. 54, Chapman & Hall, London, 1996, 277pp.
[7] B.I. Hayman, The analysis of variance of diallel tables, Biometrics 10 (2) (1954) 235–244.
[8] D.C. Hoyle, M. Rattray, PCA learning for sparse high-dimensional data, Europhys. Lett. 62 (1) (2003) 117–123.
[9] I.M. Johnstone, On the distribution of the largest eigenvalue in Principal Component Analysis, Ann. Statist. 29 (2) (2001) 295–327.
[10] R.A. Kempton, The use of biplots in interpreting variety by environment interaction, J. Agri. Sci. Cambridge 103 (1984) 123–135.
[11] J. Mandel, A new analysis of variance model for non-additive data, Technometrics 13 (1) (1971) 1–18.
[12] P. McCullagh, Invariance and factorial models (RSS discussion paper, October 13, 1999) reply, J. Roy. Statist. Soc. B 62 (2000) 209–256.

[13] M.L. Mehta, Random Matrices, 2nd ed., Academic Press, New York, 1991.
[14] S.D. Silvey, in: Statistical Inference, 3rd ed., Chapman & Hall, London, 1979, 191pp.
[15] C.A. Tracy, H. Widom, On orthogonal and symplectic matrix ensembles, Comm. Math. Phys. 177 (1996) 727–754.
[16] J.W. Tukey, One degree of freedom for non-additivity, Biometrics 5 (1949) 232–242.